From:	Stuber, Robyn
To:	Denton, Debra
Cc:	Smith, DavidW
Subject:	FW: Test Drive Input
Date:	Tuesday, March 10, 2015 8:26:00 AM
Attachments:	DMS-#3256981-v1-CASA_ATP_Letter_(Final)_pdf.PDF

From: Heil, Ann [mailto:AHeil@lacsd.org]
Sent: Monday, March 09, 2015 2:49 PM
To: Morris, Cris@Waterboards
Subject: Test Drive Input

Cris, a letter that CASA recently sent to EPA contained a nice summary of our concerns regarding the Test Drive and what it can and cannot be used to provide. I've attached the letter. The relevant portion is comment 3, which I've put below. Thanks again for being willing to think about this. - Ann

3. Comparability Data Based on the State Water Board "Test Drive" is Flawed and Cannot Serve as the Basis for Potential Future ATP Requests

As noted above, federal regulations require an ATP applicant to provide comparability data for the performance of the proposed alternative procedure compared to the performance of the approved method (40 CFR §136.5(c)(5)). The State Water Board's ATP application seemingly relied solely on the results of a TST and NOEC comparison "test drive" (State Water Board, "Effluent, Stormwater, and Ambient Toxicity Test Drive Analysis of the Test of Significant Toxicity (TST)" (December 2011).) This analysis ultimately determined that the TST identified a similar number of final effluent and receiving water toxicity tests as "toxic" as the NOEC. However, there are several flaws with this analysis.

First, the "test drive" analysis did not compare or evaluate the impact of reducing the minimum number of concentrations from five and a control to one and a control. All of the final effluent data used in the analysis were selected among valid WET tests submitted to the regulatory authorities for NPDES compliance determination. Therefore, all of the final effluent tests used to compare the NOEC and TST were obtained from tests using a minimum of five concentrations and a control that would have incorporated all protocol-required QA/QC and data validation procedures, including evaluation of the concentration-response relationship. Additionally, the "test drive" included a sizeable number of ambient/receiving water toxicity test results. All of these ambient/receiving water toxicity tests were conducted using a single concentration and control test design, and the number of tests identified as "toxic" with the TST and NOEC were also found to be similar. However, this study did not and could not evaluate and compare results from tests conducted using a five concentration and control NOEC design to those on the same samples obtained using a single concentration and control TST test design.

Furthermore, the "test drive" analysis mischaracterized these findings in claiming that the TST correctly identified more "truly toxic" or "truly nontoxic" tests than the NOEC. All of the tests were conducted on actual final effluent and receiving water/ambient samples. Therefore, the "true" or "actual" toxicity of any sample is unknown. The "test drive" erroneously inferred that if a sample exhibited a 25% effect or greater that it was "truly toxic" or if a sample exhibited an effect of 10% or less that it was "truly nontoxic." As the USEPA found in its 2001 interlaboratory validation study

using “true” nontoxic blank samples, effects as high as 80% can be observed by some laboratories when analyzing a sample that is completely nontoxic. The interlaboratory validation study determined that laboratories finding completely nontoxic blank samples “toxic” was not a rare event. Before consideration of concentration-response relationships, this study found that 15% of Ceriodaphnia reproduction tests on blank samples were incorrectly determined to be toxic and 13% of fathead minnow growth tests on blank samples were incorrectly determined to be toxic. This well documented finding would refute any conclusion that a test that exhibited a 25% effect or greater was “truly toxic.” Likewise, although not empirically quantified, it can also be assumed that actual “toxic” samples will, on some occasions, exhibit effects less than 10%. Although this was true for most endpoints, this was not the case for the fathead minnow endpoints and any comparability data submitted for any future ATP approvals should be made on an endpoint-by-endpoint basis.

A somewhat improved means of projecting the true performance of a TST test design using a single concentration and a control relative to the performance of the promulgated toxicity test design using five concentrations and a control with consideration of concentration-response would be to examine USEPA’s inter-laboratory study. In this study, for Ceriodaphnia reproduction, the number of non-toxic blank samples incorrectly exhibiting toxicity (false positives) dropped from 15% to 4% when concentration-response was considered. Similarly, for fathead minnow growth, the number of non-toxic blank samples incorrectly exhibiting toxicity dropped from 13% to 4% when concentration-response was considered. Based on this information, the lack of a concentration-response evaluation in the single concentration plus control TST method would be expected to significantly elevate the false positive error rate, perhaps as much as tripling it. The only reliable means of comparing the performance of the TST using a single concentration and a control to the performance of the TST using five concentrations and a control (or comparing the TST using a single concentration and a control to a NOEC using five concentrations and a control) would be to conduct a study using non-toxic blank samples. This should be done prior to any future ATP approval of a single concentration plus control TST method for use in NPDES permits. The “test drive” cannot be used to estimate this critical error rate, which must be determined to assess the accuracy and suitability of the test method.